14.2 A 1.9Gb/s 358mW 16-to-256 State Reconfigurable Viterbi Accelerator in 90nm CMOS

Mark Anders, Sanu Mathew, Steven Hsu, Ram Krishnamurthy, Shekhar Borkar

Intel, Hillsboro, OR

The Viterbi algorithm for decoding convolutional codes is one of the most performance and power-critical algorithms for multi-Gb/s wireline and wireless communication, where decoder specifications have a wide range of requirements for throughput, constraint length, generator polynomials and coding rate [1,2]. A 16/32/64/128/256-state coarse-grain reconfigurable Viterbi engine (Fig. 14.2.1), targeted for on-die special-purpose channel coding acceleration in high-performance processors, is fabricated in 90nm dual-V_t CMOS technology [3]. Total power consumption is 358mW, with a peak data rate of 1.9Gb/s for 32-state operation (measured at 1.3V, 50°C). Radix-4 ripple-carry add-compare-select (ACS) circuits with 1-level lookahead, double-buffered path metric (PM) memory with flexible read/write control and programmable mirrored branch metric circuits enable reconfigurable computation of ACS decision bits for a wide range of standards. Furthermore, tree-bitline traceback (TB) memory circuits with split-decoder precharge, TB read/write control/data routing with ring-buffer circuits and a reconfigurable next address generator enable energy-efficient decoding with dense layout occupying 0.53mm^2 (Fig. 14.2.7) while achieving: i) constraint lengths up to K=9 with reconfigurable generator polynomials, ii) depunctured coding rates of 1/2 or 1/3, iii) traceback lengths up to L=140, iv) data rates up to 3.8Gb/s for 16-state modes, v) 20% power reduction when operating in 16-state vs. 256-state mode, vi) fine-grained clock/supply gating enabling 24X power reduction in standby mode, vii) scalable performance up to 2.35Gb/s, 813mW measured at 1.7V, 50°C, viii) low voltage/frequency mode performance of 100Mb/s, 26mW measured at 1V and 200MHz clock.

The branch metric unit (Fig. 14.2.2) calculates Euclidean distance error metrics from up to 6 3b soft inputs (ranging from a strong '1' of 011 to a strong '0' of 100 in 2's complement format). These circuits handle coding rates up to 1/3 following depuncturing. A counter with programmable outputs produces the expected input bits for different generator polynomials. These expected inputs are XORed with their corresponding 3b soft inputs and summed to produce the 6b error metric, with a constant value added to center the range of errors around 32. This enables a mirrored design allowing one branch metric circuit to be used for two metrics with opposite expected bits, leading to a compact branch metric unit implementation occupying 0.02mm^2. This reconfigurable design results in up to 21% power reduction obtained by setting a default value to unused inputs during depuncturing and higher coding rates.

The ACS unit accumulates error metrics using 8 10b radix-4 ripple-carry ACS circuits to complete 2 radix-4 butterflies each cycle [4], chooses the minimum error at each trellis state and updates the 10b path metrics in a 256-entry PM memory (Fig. 14.2.3a). The ripple-carry nature of the ACS enables efficient reconfiguration as 8b datapath for higher coding rates by forcing the lower 2 LSBs to 0, resulting in 9% reduction in ACS power. Static configuration registers enable programmable counter outputs, providing flexible read/write control for each quarter section of the PM memory (Fig. 14.2.3b). This allows different trellis configurations to be mapped to the datapath, while using only a single read and dual write-ported double-buffered dynamic register file PM memory. The trellis butterflies are mapped such that branch metrics can be used twice due to symmetry, for true and complementary expected bits. Memory reads alternate between upper and lower memories in a sequential manner, triggered by a control signal from the counter. The PM memory is double-buffered to allow simultaneous reads from the current trellis state and writes to the next trellis state. With two ACS radix-4 butterfly circuits, computation for 2 output bits is completed in 2^{K-1} / 8 cycles. During the last cycle before starting a new symbol, the contents written to the PM memory are transferred to the read side of the memory cell. The ACS also supports one level of bypass when writing to the memory, to avoid a one cycle pipeline stall between symbols when a write/read dependency exists.

The different configurations of the Viterbi accelerator support varying memory requirements (Fig. 14.2.4a). Supply grids to the PM and TB memory circuits are therefore partitioned into 4 sections with PMOS sleep transistors inserted between the main supply rail (Vcc) and the virtual supplies (Vcc_{32}, Vcc_{128}, Vcc_{256}) to selectively shut off unused sections (Fig, 14.2.4b). Compared to 256-state mode of operation, traceback unit active power is reduced by 68%, 61% and 28% while running in the 16, 32/64 and 128-state modes, respectively. In stand-by mode, fine-grained clock and supply gating results in total Viterbi accelerator power consumption of 15mW (24× reduction).

The 40Kb TB memory is organized as two 128×160b register file (RF) sections with the trellis state stored in columns, enabling TB lengths up to L=140. 16 ACS decision bits are written to the TB memory each cycle, half to each section of memory. Write select signals $wsel_i$ are generated using column and row ring buffers that separately rotate a single '1' through shift registers (Fig. 14.2.5a). The column buffer (clocked every symbol clock) together with the row buffer (clocked every system clock) select individual 8b words for write access. The symbol clock occurs every $2^{K-1}/8$ system clocks. Write data activity is reduced by using an additional row shift register to route data to the appropriate row. This ring-buffer based write topology reduces write decoder and data power by 29% and 94% respectively over conventional RF write circuits. Separate row and column read addresses are also used to read a single bit, which is then routed using an alternating horizontal and vertical tree-bitline structure to reduce bitline wiring congestion. This organization, along with selective precharge using split-decoder circuits [5] and shared column-select transistors, results in a 90% reduction in read power compared to conventional RF designs. During traceback, the next row address is obtained using a reconfigurable tree to XOR, according to the generator polynomial, some bits of the current address and decision bits read from the memory to determine the state from the previous stage of the trellis (Fig. 14.2.5b). Column address is generated using a separate ring buffer with localized clock-gating logic that enables the clock to the 8b section that contains a '1', reducing read decoder clock power by 39%. In-order generation of decoded bits is enabled by the addition of a path memory to the TB unit, allowing address caching during traceback. For shorter constraint lengths, the write/read column ring buffer lengths can be reduced by gating the supply voltage (Vcc_{256}) to the outer sections of memory. The traceback design techniques described above result in a compact implementation occupying 0.43mm^2.

The Viterbi accelerator operates at a maximum frequency of 3.8GHz with peak throughput of 1.9Gb/s (measured at 1.3V, 50°C), consuming 358mW total power for 32-state operation (Fig. 14.2.6a). Total active leakage power component is 44mW (12% of total power). Performance is scalable up to 4.7GHz, 2.35Gb/s consuming 813mW (measured at 1.7V, 50°C). At a nominal supply of 1.3V, the accelerator achieves peak throughputs of 3.8/0.95/0.44/0.22Gb/s for 16/64/128/256-state modes, consuming a total power of 353/358/383/441mW, respectively (Fig. 14.2.6b). Operating in 16-state mode, with the shutdown of unused portions of the accelerator, reduces the total power consumption by 20%.

Acknowledgments:
The authors thank the CRL prototype team for layout help; Pyramid Probe Card Division of Cascade Microtech Inc. for high bandwidth membrane probe solution; and M. Haycock, J. Schutz, S. Pawlowski, K. Soumyanath, H. Alavi, M. Guzman, A. Chun, M. Aguirre for encouragement and support.

References:
[1] N. Bruels, et al., "A 2.8 Gb/s, 32-state, radix-4 Viterbi Decoder Add-Compare-Select Unit," *Symp. VLSI Circuits*, pp. 170-171, June, 2004.
[2] M. Bickerstaff, et al., "A Unified Turbo / Viterbi Channel Decoder for 3GPP Mobile Wireless in 0.18µm CMOS," *ISSCC Dig.Tech. Papers*, pp. 124-125, Feb., 2002.
[3] K. Kuhn, et al., "A 90nm Communication Technology Featuring SiGe HBT Transistors, RF CMOS, Precision R-L-C RF Elements and 1µm^2 6-T SRAM Cell," *IEDM Tech. Digest*, pp. 73-76, Dec., 2002.
[4] M. Anders, et al., "A 64-State 2GHz 500Mbps 40mW Viterbi Accelerator in 90nm CMOS," *Symp. VLSI Circuits*, pp. 174-175, June, 2004.
[5] S. Hsu, et al., "An 8.8GHz 198mW 16x64b 1R/1W Variation Tolerant Register File in 65nm CMOS," *ISSCC Dig. Tech. Papers*, pp. 448-449, Feb., 2006.

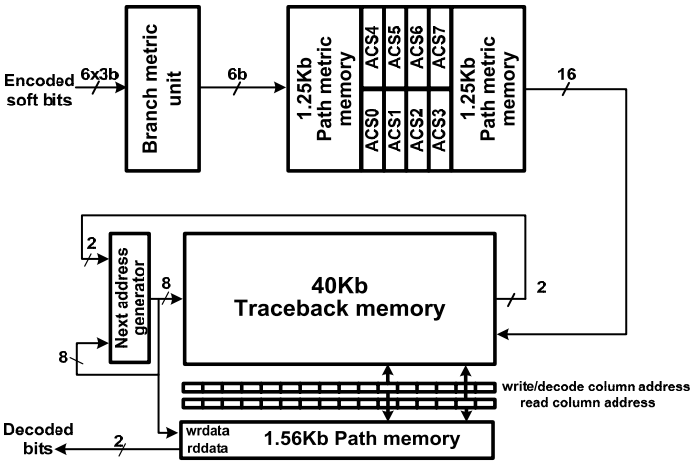


Figure 14.2.1: Viterbi accelerator organization.

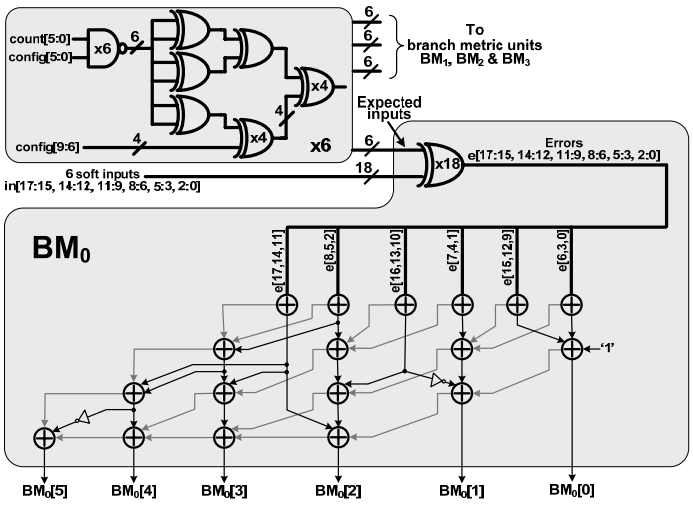


Figure 14.2.2: Branch metric circuits.

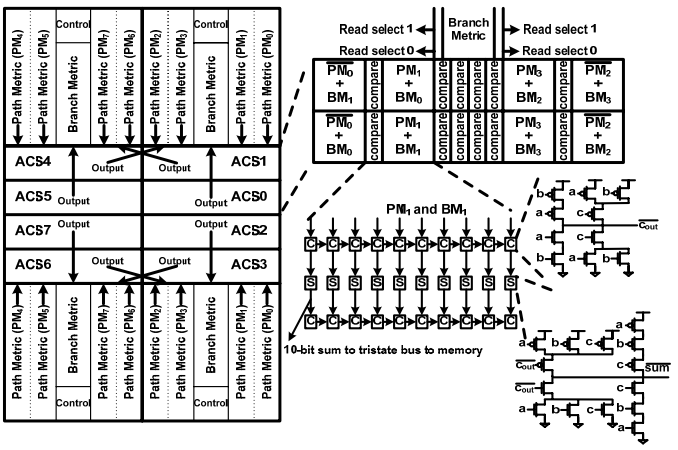


Figure 14.2.3a: Reconfigurable ACS organization.

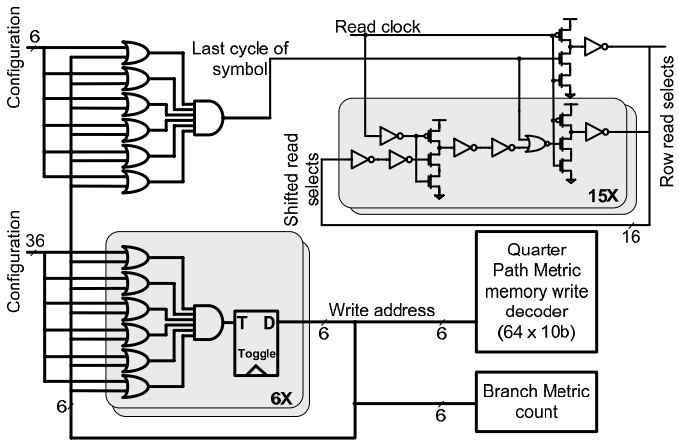


Figure 14.2.3b: Reconfigurable ACS control.

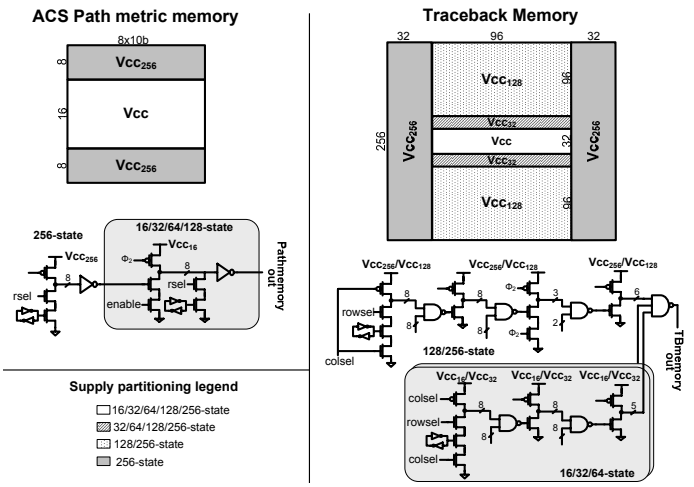


Figure 14.2.4a: Reconfigurable path metric and traceback memory partitioning and circuits.

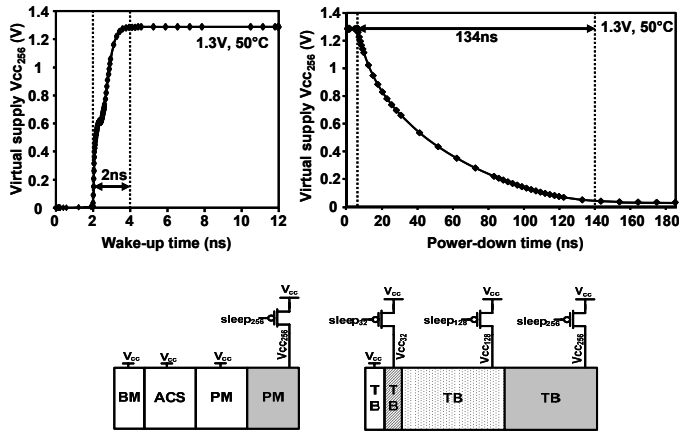


Figure 14.2.4b: Viterbi accelerator sleep transistors and virtual supply wake-up/power-down time.

Continued on Page 600

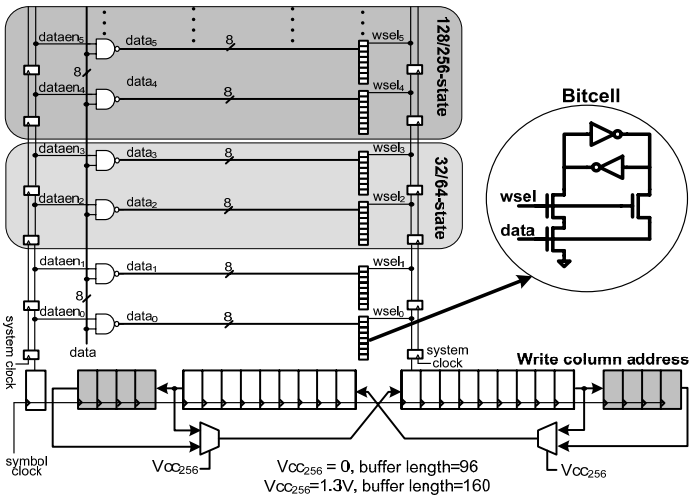


Figure 14.2.5a: Reconfigurable traceback memory write circuits.

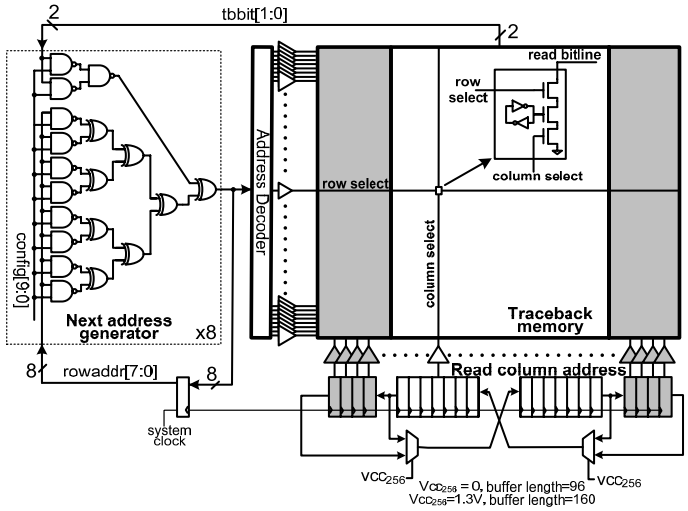


Figure 14.2.5b: Reconfigurable traceback memory read circuits.

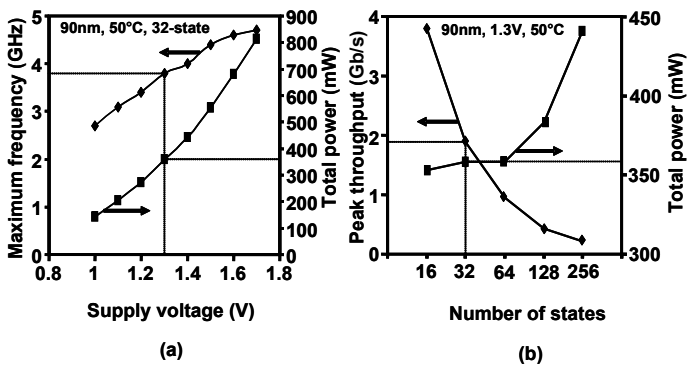
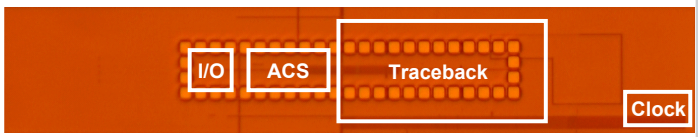


Figure 14.2.6a: Frequency/power measurements vs. supply voltage for 32-state operation b: Peak throughput/power measurements vs. number of states.



Process	90nm dual-V_t CMOS
Die area	1.9mm²
Viterbi accelerator layout area	0.53mm²
Worst-case power	358mW at 3.8GHz, 1.3V, 50°C (nominal)
Active leakage power	44mW at 1.3V, 50°C (nominal)
Nominal performance	1.9Gb/s at 1.3V, 50°C (nominal)
Peak performance	2.35Gb/s at 1.7V, 4.7GHz, 50°C
Low-voltage mode performance	100Mb/s, 26mW at 1V, 200MHz, 50°C
Stand-by mode power	15mW (24X reduction vs. active mode)

Figure 14.2.7: Viterbi accelerator die micrograph and measured performance summary.

